DOCUMENT RESUME

ED 179 613 UD 019 829

AUTHOR Pearn, M. A.
TITLE Employment Testing and the Goal of Equal Opportunity: The American Experience.
INSTITUTION Runnymede Trust, London (England).
PUB DATE Jul 78
NOTE 40p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Affirmative Action; Aptitude Tests; Blacks; Disadvantaged Groups; *Equal Opportunities (Jobs); *Minority Groups; *Occupational Tests; Racial Discrimination; Standardized Tests; Supreme Court Litigation; *Test Bias; *Test Validity

ABSTRACT

Testing in the United States has had a checkered history since tests were developed and used at the beginning of this century. In the earlier part of the century, test results were used to justify the exclusion of immigrants from parts of Europe. More recently, they have come under attack as restricting the access of blacks to jobs and education. Following the Civil Rights Act of 1964, there was an increase in litigation about testing, and as an alternative to employment tests, many employers instituted special programs and practices to insure minority hiring and advancement. The past few years have seen a greater flexibility on the part of the courts and an increase in the use of employment tests, with greater attention paid to the question of test validity. Testing is now on a firmer footing and should not be judged by the inadequacies of the past. With due regard to the reduction of bias in their use, validity used job related tests can contribute to the removal of arbitrary barriers to the employment of minority workers. (Author/RLV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179613

The Runnymede Trust

62 Chandos Place
London WC2N 4HG
Telephone 01-836 3266

EMPLOYMENT TESTING

AND THE GOAL OF EQUAL OPPORTUNITY:

THE AMERICAN EXPERIENCE

By

Dr. M.A. Pearn - Deputy Director

CONTENTS

	Page
Introduction	1
Employment Testing	2
Guidelines on Selection	27
Conclusions	33
References and Notes	35

July 1978

Price: £1.25

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OF FICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Pearn
Runnymede Trust

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

UD 019 829

Employment Testing and the Goal of Equal Opportunity :

The American Experience

M.A. Pearn

Introduction

Tests for selecting employees have long been used in Britain, though they are more likely to be found in certain industries, and to be used for certain job categores.

In their own interest employers using selection tests, or other selection procedures, should use them as effectively as possible, though there is no law in this country which requires them to do so. However the definition of indirect discrimination in the Race Relations Act 1976 could require an employer to justify his selection practices if a complainant could prove to an industrial tribunal that it put him at a disadvantage because of his or her colour, race, national or ethnic origin, or nationality. In other words, if he or she could satisfy an industrial tribunal that the selection procedure rejected a higher proportion of members of a particular racial group, then the employer would be required to show that the practice is justifiable on non-racial grounds. This can be a complex technical matter involving validation strategies, job analysis, measures of worker performance, and complex statistics, especially where tests are used. If the employer failed to justify the practice, it would be declared unlawful, and he could be ordered by the tribunal to refrain from using the procedure in question. Exactly the same conditions apply under the Sex Discrimination Act 1975 which has an equivalent definition of indirect discrimination.

This is a new problem for both employers and test experts in this country. In addition it is unclear, until cases start accumulating, just what a tribunal might accept as the justification for the use of a selection procedure which rejected a higher proportion of black applicants. The United States with similar legislation has been grappling with this problem for many years, particularly since 1971 when the U.S. Supreme Court ruled on its first employment case under the Civil Rights Act 1964.

The landmark decision in the case of Griggs v. Duke Power Co. has had a direct influence on the formulation of the definition of indirect discrimination in the Sex Discrimination Act 1975 and the parallel definition in the Race Relations Act 1976. Consequently, the American debates, in and out of court, can be of great value in the UK in creating an awareness of the issues involved, and perhaps even clarifying them.

In addition a great deal of research has been generated in the United States on ways of overcoming, or adjusting for, racial and cultural bias in selection, which was directly stimulated by the Civil Rights legislation. There could be many practical lessons for Britain.

Because of the potential relevance of U.S. experience, the author felt that a close look at recent U.S. developments would be productive. An application was made to the Ford Foundation in New York for funds to enable the author to visit the United States to talk to employers, researchers, academics, civil rights law enforcement agencies, and civil rights organisations. The Ford Foundation granted the author a study and travel award for one month. The Runnymede Trust thanks the Ford Foundation for their help and sponsorship, which has been forthcoming on two previous occasions.*

This report is an attempt to distil some insights and practical lessons from the myriad impressions gained from the visit, from the literature, and from the many hours of discussion in the United States with people whose enthusiastic interest and helpfulness were of the greatest value.

It is not possible to examine testing and unbiassed selection procedures without setting the topic firmly in the broad context of equal employment opportunity in the United States. This has already been described in detail by Tom Rees. This report in part updates the earlier paper, but concentrates mainly on the developments in employment testing and equal opportunity.

I Employment Testing

Testing in Britain and the USA

Although the main emphasis of this paper is on personnel selection tests, (viz. measures of aptitude, personality or general ability) there is a wider sense in which tests can be defined as any measure on which a promotion or selection decision is based. This could include, among other things, the use of interviews, application forms, and ratings if they are scored on a formal or standardised basis. Standardised tests form an important part of American personnel practice, and the publication and distribution of tests has grown into big business, with millions of tests sold every year.

* See the reports, following their visits to the United States, by David Stephen "Minority Rights and Minority Morale in the USA - How Relevant is U.S. Experience?" May 1973, and by Tom Rees, "Equal Employment Opportunity in the USA - Recent Developments", August 1974. Both reports are available from the Runnymede Trust.

By contrast, in this country, though widely used during the War, standardised tests have never caught on in a big way, though they tend to be concentrated in certain industries, and used for certain kinds of jobs or training.[1] They are commonly used in the public rather than the private sector, and the government itself, through its agencies, is a big user of tests.

Testing in the United States has had a chequered history since tests were first developed and used at the beginning of the century, after their importation from Europe. Some of the early pioneers of mental testing in the United States used the supposedly scientific and objective data from their tests to support and justify the Immigration Act of 1924 which had the effect of excluding migrants from parts of Europe whom the testers had shown to be 'biologically inferior' to those coming from other parts.[2] More recently, they form the basis of the controversy over the heritability of intelligence and the postulated difference in the average levels of intelligence between racial groups.[3] In terms of the effects that education and employment tests have on racial equality, they can be likened to the old 'Jim Crow' laws of the Southern States. These were unwritten rules in the Southern States which governed the places to which Negroes could have access. Now it is argued that selection tests in employment (and those used for educational purposes) are having the same effect in restricting the entry of blacks to jobs and education. This is because blacks frequently obtain lower scores on tests and, as a result, higher proportions fail to obtain jobs and training by comparison with the white population. A critical issue is whether the tests are as fair a measure of the abilities and potential of blacks as they are for whites. If not, can alternative measures or procedures be found?

On the positive side, tests played a critical role in the United States during the first and second World Wars when millions of people were tested in order to assign them to military or industrial jobs. Since then testing has never looked back, and the importance of testing in American culture is reflected in the special reference to testing, known as the Tower Amendment, in the Civil Rights Act 1964 : *

* The Civil Rights Act 1964 protects from discrimination racial groups, religious groups, and women. Under other legislation, people between the ages of 40 and 65, and the physically handicapped are also protected. This leaves only white physically able males under the age of 40 who are not protected by anti-discrimination laws. Some employers I spoke to were so anxious to find employees in the protected classes, so as to avoid charges of discrimination, that they claimed they were no longer recruiting young white males.

> "Nor shall it be an unlawful employment practice for an employer to give and act on the results of any professionally developed ability test provided that such test, its administration, or action upon the results is not designed, intended, or used to discriminate because of race, colour, religion, sex, or national origin." (Section 703(h))

The Race Relations Act 1976 does not specifically mention tests but they can be subsumed in the definition of indirect discrimination in the Act as a general condition or requirement imposed by an employer. In the Race Relations Act, indirect discrimination is defined as the general application of a condition or requirement which is such that the proportion of persons of a racial group able to comply with it is considerably smaller than the proportion of persons not of that racial group who are able to comply, and which the employer cannot justify on non-racial grounds.[4] Indirect discrimination is defined in terms of effects, rather than motives, and the provision is designed to prohibit practices which may appear fair in a formal sense but which are discriminatory in their effects. Under this provision, if an employer can justify the requirement or condition, then it is not unlawful, even though it may result in a substantially higher rejection rate for racial minorities.

It is interesting to note that the use of tests was mentioned in the Government White Paper[5], preceding the Race Relations Bill, as an example of a possible source of indirect discrimination, though it should be remembered that educational or work experience requirements, proficiency in English, and the use of biographical information could also come under this heading.

The Organisation of Equal Employment Opportunity Effort in the United States

The main government agency in the United States enforcing the Civil Rights employment provisions (known as Title VII) is the Equal Employment Opportunity Commission. Until the Civil Rights Act was amended in 1972 the EEOC was little more than a conciliation agency concerned only with private employers. Under the Equal Employment Opportunity Act 1972, its scope and powers were increased. The EEOC is a powerful and energetic organisation, and it is also controversial. Its scope includes state and local government, but not the Federal government which polices its own employment activities internally, through the Civil Service Commission.

The third agency in the field is the Department of Labor which is responsible for the Federal Contract Compliance Programme. Under separate legal authority, (viz. presidential Executive Orders[6]), any employer holding, or applying for, a government contract must comply with regulations which are designed to ensure that they eliminate discrimination, and that in addition they take active steps to promote equality of opportunity through an affirmative action programme. The conflict of interests and the disagreements between these three Government agencies reveal and highlight issues of relevance in Britain. These will be discussed at a later stage in this report. There are now moves under President Carter's Civil Rights Re-Organisation Program announced in 1978, to widen the scope of the EEOC still further, to include duties carried out by the US Civil Service Commission and some by the Departments of Labor and Justice.

Employers and Equal Opportunity

A visitor from Britain cannot fail to be impressed by the sheer magnitude and complexity of the government machinery that has been set up to eliminate discrimination, and to create equal opportunity for all. It is clear that progress has been made and that there is evidence of some upward mobility among blacks. A recent survey conducted by the Bureau of National Affairs examined the equal employment opportunity effort of 160 organisations.[7]

This survey showed that active measures to prevent discrimination and encourage equal opportunity are firmly established as "a fact of life" by nearly all employers, and that the personnel function has changed in a variety of ways as a result of the Federal Government's effort to enforce the Civil Rights Act. This contrasts sharply with the results of similar surveys in Britain where most employers are characterised by apathy and uninterest.[8] The majority of employers in the American survey indicated that they had made changes in their selection procedures, especially with regard to testing, and had revised job qualifications. Other measures included special recruitment and training programmes, similar to those now permitted under sections 37 and 38 of the Race Relations Act.* One measure of the effects of these changes was the finding that 49% of the companies surveyed reported blacks in management jobs previously held only by whites.

* Under this provision, wherever racial minorities are under-represented in jobs, either nationally or locally, an employer or other specified bodies, can provide special training for the under-represented racial group. The employer can provide special encouragement to the racial groups to apply for the jobs in question, but cannot give them preference in selection. The aim of the training would be to raise black workers to a more competitive level of employability by equipping them with necessary job skills.

Although this could be taken as a measure of progress the numbers fall far short of equal representation by blacks in the labour force. Although the overall average percent of black workers in local labour markets is 17.3%, only 5% of officials and managers, and only 3% of sales staff, in the survey were black. On average the total workforce in the companies surveyed tended to have a proportionate number of black employees, but their distribution in the organisational hierarchy was very uneven, usually clustering in the lower level jobs. This pattern varied according to the type of industry and the size of the company:
Manufacturing and small companies were worst, whilst the non-business sector showed the best distribution of black workers. One reflection of the sheer level of the government enforcement effort was the finding that 63% of the employers in the survey had been subject to equal opportunity investigations or other legal actions, half of which were upheld against the company. Among these companies who were subject to Federal Contract Compliance inspections, a third of those reviewed were found to be not in compliance.

In talking to employers* the sense of commitment to equal opportunity becomes very apparent. There is no doubt that most employers take the civil rights enforcement efforts seriously, even though they may not always agree with the conditions imposed. Some employers seem to take a very defensive stand and have full-time staff doing 'vulnerability analysis' in order to ensure that there is no case for legal action against them. Most employers are ready to admit that the main motivator has been fear of litigation, and the severe penalities which the courts have imposed, though there is a general feeling that the attitude of the courts is beginning to soften, and that they are now showing greater sympathy for the employers' position. It is alleged that, in the early days of the Civil Rights Act, employers nearly always lost a case involving testing. Recently, and particularly in 1977 the courts appear to be taking a more flexible line in testing cases. In particular, a recent Supreme Court case involving testing has shown less deference to the stringent standards embodied in the Selection Guidelines issued by the Equal Employment Opportunity Commission. It is also widely believed that the $38m settlement in the AT & T consent decree in 1971 could never be repeated.[9] The increased passage of time since the race riots of the sixties, and the associated decline in white guilt and anxiety, the worsening economic position, and a less sympathetic ear from the courts make the general climate less urgent. One reason given for the changing attitudes of the courts is that recent Administrations have been less interested in Civil Rights.

* The auther visited Citibank, CBS, Rockwell International, AT & T, Sears Robuck, and talked to representatives of many others.

President Johnson was careful to appoint judges who were informed and enlightened on race relations. Since then there has been a marked decline in successive Administrations' interest and this has been reflected in judicial appointments. Another factor of increasing importance is that the civil rights movement is beginning to lose out to competition from the Womens Movement which is gaining ground, and sometimes at the expense of opportunities for the ethnic minorities.

This apparent easing up of the presssure toward equal opportunity cannot be attributed solely to its success in the past, as America is still a divided and unequal society with an unemployment rate of 14% among blacks which is twice that for whites. The National Urban League's hidden employment index actually puts the figure much higher (at 35%), if the underemployed are also taken into account. Even the official unemployment rate for blacks, aged 16-19 years, is over 40%.[10]

In addition, the evidence of racial disadvantage and inequality is enormous.[11] While the number of blacks earning more than $10,000 p.a. is increasing, and those earning between $5,500 and $10,000 is shrinking, almost a third of all black families are still below the poverty line, defined as $5,500 for an urban family of four, whereas only 9% of white families are below the line. Although many blacks have risen to the secure middle class in recent years, the substantial majority of America's blacks remain in the "underclass",[12] crammed into the declining inner cities without hope of employment, and beset by poor housing, inadequate social services, and physical insecurity. Some blacks have benefited from the legislation, but it seems to have by-passed the most disadvantaged.

Testing and the Civil Rights Act

It is in this context that employment tests are seen by the civil rights lobby as an obvious means by which blacks are excluded from jobs, and it is not surprising that tests quickly became the focus of complaints alleging discrimination by employers, and the subject of many important court cases. The initial effect of the court cases was to expose the prevalence of extremely poor testing practices such as the use of a test in the absence of any validation data, or with arbitrarily set pass marks, which in some cases was a deliberate ploy to discriminate against blacks. With the advent of the EEOC Selection Guidelines and their endorsement in the courts, many employers became nervous of using tests. The EEOC Guidelines (which are discussed in detail below) imposed stringent requirements on the employer to validate a test if its use resulted in the rejection of a higher proportion of black workers. Many employers,out of fear of litigation abandoned the use of tests altegether, or now use them in a more flexible way, (e.g. without fixed cut-off scores). There has however been a recent resurgence in the use of tests,

which may in part be a response to more lenient attitudes taken by courts. Another effect of the litigation on the use of tests has been the move toward greater sophistication in the way tests can be used, and a greater awareness of the role they can play in personnel selection. This move is closely related to the opening up of the debate on the fairness of selection practices, and a fresh examination of old ideas about the validity of tests. Linked with this debate is a surge of new research into cultural bias in selection and the development of culturally fair techniques and procedures. These issues and the research will be discussed below but first it may be useful to examine the critical role played by the United States Supreme Court.

Personnel Testing and the Supreme Court*

At the highest court in the land, the Supreme Court only considers issues on which there have been conflicting ruling by different appeal courts. It was not until nearly six years after the Civil Rights Act became effective (in 1965) that the United States Supreme Court considered its first employment case. Until that time the lower courts had built up a considerable body of case law and had established a basic framework of interpretation of the 1964 Civil Rights Act. It was in the case of Griggs v. Duke Power Co. that the Supreme Court ruled that it is not necessary for an employer to intend to discriminate in order to violate the law.[14] Another issue in the case was the use of the Wonderlic and Bennet Mechanical aptitude tests, and it is of value to trace the history of the case.

Prior to the effective date of the Civil Rights Act, the Duke Power Co. had only employed blacks as common labourers, but from that date it introduced the tests, setting cut-off scores based on the national average for high school graduates, with no attempt to evaluate the effectiveness of the test, or to interpret it by reference to the local population.** In addition the employer had for some time required high school diplomas for any job, other than labourer. The combined effect of the educational and test requirements was to screen out proportionately more black employees than white employees. Lower courts considering the case had ruled that lack of discriminatory intent in the use of the tests by the employers was a defence, whereas in other cases the courts had ruled that it was not a defence. Finally, the Supreme Court decided to resolve the issue. It is worth quoting at length from the unanimous opinion of the court, as it is instructive in elucidating the meaning of indirect discrimination which the Race Relations Act 1976 derived from it:

* These accounts of Supreme Court decisions rest heavily on discussions with, and papers by, Thaddeus Holt, attorney for the American Society of Personnel Administration.[15] In addition, further material came from Dee-Ann horstman of the Civil Service Commission and from her excellent summary of Court cases in : "1975-1976 Court Case Compendium : Legal Standards for Personnel Practices." Professional Series 77-5. Washington, DC : U.S. Civil Service Commission.

** In other words, the employer made no attempt to validate the use of the test.

"Under the (Civil Rights) Act practices, procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be operated to 'freeze' the status quo of prior discriminatory employment practices".

"The Act prescribes not only overt discrimination but also practices which are fair in form but discriminating in operation. The touch-stone is business-necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to proper performance, the practice is prohibited.

"Good intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that are unrelated to measuring job capability".

This ruling imposed on the employer the obligation to present evidence that the employment practice was job-related, once the complainant had established a prima facie case of discriminatory effect. Unfortunately the question of what constitutes the job-relatedness of the selection criteria was left unanswered, as the employer in the Griggs case presented no evidence of job-relatedness at all. Under this ruling, it was not sufficient that the tests used were "professionally developed", as stressed in the Tower Amendment ; their use should have a demonstrable relationship to the jobs for which the applicants were assessed.

It was not until 1975 that the Supreme Court took the opportunity to address these issues again, in the case of Albermarle Paper Co. v. Moody.* One of the two issues before the court was the question of what constitutes a sufficient showing of job-relatedness, with respect to an aptitude test. For the first time the technical complexities of test validation were brought before the Supreme Court, and raised the possibility of the justices trying to resolve technicalities which were inadequately dealt with by the lower courts, the practical implications of which would have enormous impact on employers' testing practices. The Albermarle Paper Co. had in fact conducted a validation study, but only some time after the litigation had started years earlier in a lower court. The validation study was assessed by the court in the light of the EEOC's Guidelines on Selection in a way which appeared to give them the force of law, instead of the advisory status defined by the legislation.

* The Supreme Court only hears a small proportion of the cases submitted to it.

The validation study, which was described as an "odd patchwork of results", was faulted on four counts :

(a) The study revealed that the tests related only to three of the eight lines of job promotion for which they were used. In other words performance on the test predicted performance on only some of the jobs. Following the Guidelines, the court declard that a test may be used in jobs other than those for which it has been professionally validated only if there are no significant differences between the studied and the unstudied jobs. The validation study had not analysed the skills or attributes required for each job. This was a clear indication by the court that the tests should be empirically related to the job.

(b) In the validation study job performance had been appraised using subjective supervisory ratings so that the actual criteria being used by the supervisors could not be established. Consequently the standard being applied was too vague, and was subject to bias and distortion. The court required that measures of job performance should be objective and free from bias.

(c) The validation study had focussed on the degree to which test scores could predict successful job performance, but had only focussed on jobs near the top of a promotional line in the validation study. The court argued that tests which predicted performance on a higher level job would not necessarily efficiently select for the lower level positions, particularly if only a small number of employees reached the higher level jobs, or there was a long passage of time between entry and the attainment of the higher level jobs. In other words, tests should be used to predict performance on actual entry-level jobs rather than possible jobs, if few employees progressed to those jobs.

(d) Finally, the court criticised the validation study for not taking into account what has come to be known as differential validity. This is the possibility that a test may be effective (defined in terms of correlations between the test and the criterion) with one group (e.g. white males) but not with another (e.g. black males). The study had utilised a strategy which basically compares the test scores of successful established workers with those who are less successful, usually defined on the basis of an assessment of their job performance.

If the test shows good differentiation between 'good' and 'poor' workers it can then be used as a basis for selection. (This is known as the concurrent model of validation.) The Albermarle validation study had been based on white experienced workers whereas the test was administered to applicants who were mainly younger, inexperienced and non-white.

In Britain, these issues will inevitably enter into considerations of what constitutes an employer's defence of a selection practice which has discriminatory effect, or to use the equivalent American phrase, "adverse impact".

Shortly after, in 1976, the Supreme Court considered its third case involving employment testing. This case (Washington v. Davis) also dealt with the justification of a selection practice which has a discriminatory effect (adverse impact). The case concerned the hiring of police officers, in the District of Columbia, particularly on the basis of a written test developed by the US Civil Service Commission and used nationwide for selection by government agencies. The test had a high degree of adverse impact, but the Supreme Court argued that the test was a useful predictor of performance in the police training academy. This, according to the court, was sufficient to prove that its use was not unlawful, despite the fact that four times as many blacks as whites failed the test. The basis of the court's decision was its acceptance that "some minimal verbal and communicative skill would be very useful, if not essential, to satisfactory progress in the training programme". This view accepted the use of success in training as a criterion for determining the effectiveness of a selection device, even though the relationship between training performance and actual job performance had not been demonstrated, and may in fact have been poor. In the Washington case the court appeared to accept a rather superficial rationalisation as a sufficient basis for establishing the job-relatedness of the test, in contrast to the empirical demonstration required in the Albermarle case.*

A very significant statement made in the opinion to the Washington case was tucked away in a foot-note: "It appears beyond doubt by now that there is no single method for appropriately validating employment tests for their relationship to job performance". This represented a departure from the previous reliance of the Supreme Court on the EEOC Selection Guidelines which futher complicated the wranglings between the different Federal enforcement agencies who took opposing views on the issue.

* The issues in the case were in fact much more complex. The Washington case was being considered under the Fourteenth Amendment of the Constitution, and the court ruled that in constitutional cases it is necessary to demonstrate discriminatory <u>intent</u>. However, statutory cases taken under the Civil Rights Act do not, following the Griggs ruling, require the proof of intent to discriminate. In this sense the Washington ruling does not represent a reversal of the earlier Supreme Court decisions, though this appears to have caused confusion among many employers. The court also argued that constitutional cases do not require the same rigorous demonstration of job-relatedness which is required under the Civil Rights Act.

The decision caused a considerable outcry among Civil Rights groups, as it appeared to represent a distinct softening of attitude towards testing. In particular, the ruling indicated that the Civil Service Commission need not comply with the stringent EEOC Guidelines on selection. The court had accepted a testing practice with high adverse impact, on arguments, rather than a demonstration, of the validity of the test. However it should be pointed out the Court did take into account the employer's successful equal employment opportunity programme which had resulted in a high proportion of blacks in the police-force. Even so a test with a high discriminatory effect was justified on an argument for, rather than a demonstration of, its relevance to training, and to the job.

In addition to these issues, lower courts have discussed and frequently ruled on the discriminatory effect of, among other things, assessment centres, educational qualifications as selection criteria, height and weight requirements, interviews, language requirements, sample sizes, sample differences, the construction of tests and job performance appraisal. Unfortunately, expert opinion on these matters (frequently industrial psychologists defending or attacking validation studies in court) reveals little consensus, leaving the courts little option but to resolve many of the issues themselves.[15] These issues will inevitably be raised before industrial tribunals under both the Race Relations Act and the Sex Discrimination Act.*

Fairness in Selection: Theory and Research

Many of the issues revolve around whether or not tests are culturally biased, but it is clear that the really important question is whether the test is, or can be, used fairly for selecting employees from different ethnic groups. The answer to that question, if it is to be accurate has to be complex. In a series of papers, R.L. Flaugher[16] at the Educational Testing Service in Princeton, N.J., has presented and elucidated many different definitions or conceptions of test bias, some of which are considered below.

* The Runnymede Trust has formed, with the British Psychological Society, a Joint Working Party on Employment Assessment and Racial Discrimination. The group is currently working on an advisory document for employers which should help clarify many of the issues, and provide guidance to test users, and Industrial Tribunals. For a brief account of the work of the group see Personnel Management, June 1978.

A major source of unfairness in selection lies in the interpretation that is put on a test score. The classic distinction between an aptitude test and an achievement test is well known. The latter measures accomplishment, whereas the former measures capacity for accomplishment. They are, of course, closely interrelated, but as Flaugher argues, there is a tendency to read a low score on an achievement test as a sign of low capacity, whereas in practice low attainment could result from poor schooling and a poor social environment. This would also apply to the use of school qualifications for selection purposes. Poor school achievement may not correlate with poor performance on the job.

Equally it should be remembered that there is no pure measure of capacity independent of achievement, which is so readily restricted by a disadvantaged background. If a group suffers from great socio-economic inequality and the effects of centuries of discrimination its average level of attainment, as measured in part by tests, will be lower than for the population as a whole. A low score does not necessarily mean low potential, though for any individual a low score could reflect low attainment and low potential.

A further source of bias, according to Flaugher, is the over-interpretation which tests can lend themselves to. He refers to the "seemingly advantageous characteristic that tests have, namely, the capacity to instill in interpreters the belief that they are assessing far more than they actually are, and what is often something different, far more than what they have claimed to be assessing". It has been pointed out that tests are frequently named, not for the kind of task presented in the test, but for what they were intended to measure.[17] A test of commonly encountered problems will not be called that, but rather a test of Practical Judgement. If suggestions by the testee are called for, then the word "creativity" will probably appear in the title. There is a large interpretative leap from an individual being unable to solve a few problems on a pencil-and paper test to his being declared lacking in practical judgement. This raises questions about the fairness of the test as an assessment device. It is not sufficient for an employer to claim that a job requires mechanical aptitude, as a justification for using a standardised mechanical aptitude test. It is clear that in employment selection the choice and use of a test should have an empirically demonstrated relationship to the jobs for which applicants are being selected.

A similar problem arises in the claim that test bias is reflected in differences in the average test scores obtained by different ethnic groups. There are many tests in which blacks and Hispanic Americans, on average, obtain lower scores than whites.[18] In the United States this is mainly a reflection of the enormous socio-economic inequality that exists between the main ethnic groups, and until the statistical arguments for racial differences in intelligence become stronger, and while the gross social inequalities persist, it is not possible to give much credence to the genetic hypothesis.*

In Britain the evidence for racial disadvantage is now considerable, though differences between ethnic groups in average test scores may also reflect cultural differences, and unfamiliarity with the home culture of the test, and the whole culture of testing itself. The main cultural groups in Britain are either recent immigrants from the West Indies, or Asians from the Sub-Continent or Africa, or their descendents (about 40% of whom were born in Britain). The immigrants may suffer from the disadvantage of being so unfamiliar with testing that they cannot hope to perform at their best, by comparison with the white British population for whom the tests were originally developed. This problem has long been known in cross-cultural studies but has not been seriously faced by British employers. The ways in which cultural differences can suppress the performance on selection tests of ethnic minority workers in Britain, has been discussed by the author elsewhere.[19]

A neat illustration of the same process working in reverse is the BITCH (the Black Intelligence Test of Cultural Homogeneity).[20] It was designed to be a culture-specific test which taps and reflects the cultural life of blacks living in the United States. Though still a development tool, it has one consistent result, viz. that, on average, blacks consistenly score higher than whites.

One strategy to overcome the cultural handicap which has been examined in the United States, is the attempt to compensate for lack of "test sophistication", as it is called, among the disadvantaged by using specially developed test pre-orientation exercises. These are designed to familiarise the applicant with the purpose of testing, to help him understand instructions, and what is expected of him in a test.

* Kamin (1977) contains a thorough critical examination of the statistical evidence for the genetic determination of racial differences in intelligence, and concludes that the evidence is not inconsistent with environmental interpretations.[2]

It sometimes gives one or two practice items, so that he can get used to the layout of the test, and the way in which he or she is to record the answers, particularly if they are multiple-choice items with a separate answer-sheet. The object of the exercise is not to give the "test-unsophisticated" person an unfair advantage but to remove some of the factors which unfairly restrict his performance on a test but not that of a person who is already familiar with testing.

The Educational Testing Service has extensively investigated the factors contributing to test sophistication and the effects of time limits on test performance. This work is no longer continued on a large scale, as the benefits were small in relation to the research effort involved. In Britain, the cultural gap between the host society and the immigrant cultures is very much greater, and this approach should be investigated particularly with newly or recently arrived immigrants who may have no prior experience of testing at all. Ideally, pre-orientation exercises should result in a gain in performance on subsequent tests which would be greatest for those from disadvantaged backgrounds, and for those from different cultures, and smallest for those already familiar with testing and the cultural values embodied in the tests. Standard packages are issued by the U.S. Employment Service*, and also by the Psychological Corporation, the major publisher of tests used by employers. There would be some value in assessing the effectiveness of these pre-orientation exercises in Britain.

Another attempt to reduce mean differences in test scores between ethnic Minority groups is the use of so-called culture-free or culture-fair tests. The main conclusion of an earlier review by the author, based largely on American research, is that there is no such thing as a culture-fair test.[19] Every test reflects and must be used in a cultural context; it is better to accept this fact and use the test accordingly. There is even some evidence that with some tests blacks actually obtain lower, and whites obtain higher, scores on culture-fair tests, widening rather than closing the gap in average scores.[21] Unless it can be demonstrated that a culture-fair test actually predicts job performance efficiently for the main ethnic groups, it is not an automatice solution to replace an existing test with a culture-fair test.

* Booklets in use include "Pre-testing Orientation on the Purpose of Testing", "Pre-testing Orientation Exercises", "Doing Your Best on Aptitude Tests".

An issue, which has generated a great deal of research and debate is the issue of differential validity (viz. whether the same test is as strong by related to job performance for each of the different sub-groups, usually defined by sex or race). In other words, is the relationship between test performance and subsequent job performance as strong for each of the different sub-groups in a given context? The issue unavoidably involves complex statistics. A large number of research studies indicate that tests invariably behave differently for different sub-groups, and a large body of research studies conclude that this finding is a statistical artefact. The issue is not just theoretical as the EEOC Selection Guidelines rest on the former assumption, even though a lively debate is still being conducted in the learned journals.[22] A six-year study by the Educational Testing Service concluded that employment aptitude tests do not behave differently with different sub-groups.[23] This leaves open the possibility that other types of tests, and all tests used in other contexts (e.g. education), may not be free from this effect. Even if demonstrated, it would be theoretically difficult to explain how the factor of race could alter the nature of the relationship between test performance and job performance, rather than just the levels of performance on the test or the job.

Flaugher takes the view that if the phenomenon of differential validity is so difficult to demonstrate, and the evidence is so debatable then it is not likely to be a very potent source of test bias, and that it is better to direct our energies to more fruitful lines of research. From a practical point of view, differences in the statistical relationships (viz. correlation co-efficients) between the test and the job are not in themselves sufficient evidence of bias in the use of the tests.

The potential source of bias lies in the way in which a test is used. It may be possible with a job-related test to reduce the adverse impact by compensating for cultural differences, lack of test sophistication, etc., with the result that the test is equally efficient in predicting good workers, regardless of race or ethnic origin. It makes sense for an employer to examine the impact a testing procedure has on the main ethnic (and sex) groups involved. The first stage would be to examine the different rates of rejection of the main ethnic groups by the test, and then to determine whether the test predicts job performance equally well for each of the main ethnic groups. It is at this point that the test user can decide whether the test is being used fairly. Consequently, the research and the associated debate whether differential validity is a general phenomenon is irrelevant from the point of view of the employer, as he should check in each instance on the possibility of unfair discriminatory effect.

Another source of bias lies in the design of the test itself, both in terms of its content, and the way the test is constructed. One of the early assumptions about test fairness was that if a test was properly (i.e. professionally) constructed then it would adequately meet the needs of the user and would be free from bias. Recently it has become clear that the main source of bias in a test, if any, results from the interpretation that is put on the test scores, and the way in which it is used. However, there are ways in which the design of the test itself can reduce the risk of unwitting biases in interpretation and use. In particular, typical or average scores (norms) obtained in the preparation of the test and usually presented in the test manual, should be based on both majority groups and minority groups. This is sensible enough but once again the issue is more complex. If, for example, the standardisation is based on a sample of Negroes in Iowa, the norms will not reflect the characteristics of the relatively more disadvantaged Negroes who live in the inner cities of southern states. Greater care can be exercised in the selection and presentation of the test content, and at least one test distributor has panels representing ethnic minorities who scrutinise items for new test for potential bias, e.g. that they may reflect knowledge that is not equally available to all groups.

An interesting development, in the field of test construction is tailored testing,[24] which has been brought about by recent theoretical developments, and through steady advances in computer technology on which it relies. In some ways the techniques function in a way similar to adaptive Programmed Instruction. The person being tested sits in front of a computer terminal screen on which a single test item appears, and he keys in the answer. Depending on whether the answer is correct or not, the next item to appear will either be easier or more difficult. With each answer the computer revises its estimate of the individual's ability which progressively becomes more stable, and the test continues until a specified level of stability is reached. There would appear to be many advantages to this technique which include considerably shorter testing time, and the possibility that cultural bias in each item can be statistically controlled for each individual as he is tested. The possibilities are very exciting but it may be some time before the technique is widely used and it does depend on expensive computer facilities, and apart from very large employers, will probably be used on a co-operative basis, with several employers sharing the same resources.

The testers' dilemma; valid tests and the goal of equal opportunity

Unfortunately the valid use of tests to overcome discrimination is rarely straight-forward. Wherever a test is used there may or may not be significant differences between the average scores of the main ethnic groups; differences in the strength of statistical relationships between test and job; and not infrequently differences between the sub-groups in their performance on the criterion (i.e. some measure of performance on the jobs for which applicants are being assessed. The practical consequences of these permutations will depend on the particular model of selection fairness which is adopted. Regretably this is another arena in which there is a raging debate which to the outsider may look like academic pillow-fighting but it is of very great practical significance. Petersen and Novick at the University of Iowa have thoroughly examined the competing models of selection fairness and compared their practical effects.[25] Perhaps the most interesting outcome of the debate over the competing models is that, although they are all objective mathematical models defining the fair use of a test or selection device, they all, with some exceptions, require and embody the adoption of certain values by the user. The values involved are the judged desirability of the practical outcomes of using the different models, e.g. discriminating unfairly against disadvantaged whites in the interests of not discriminating against any blacks who are the present victims of centuries of discrimination in American society. Another value judgement which influences the choice of model is whether group rights are more important than individual rights. It is interesting that there is no objective basis for choosing between the different models. The main basis for choosing between the different models is to determine whether or not the model has the effects the user wants it to have. In many cases this would mean using the test in such a way that a pre-determined proportion of black applicants are accepted, who might otherwise be rejected by the test if it were used without regard to the needs of minority groups.

The reason why values are important, in what is otherwise a mathematical formulation of models of selection fairness, is that there is a conflict of interests confronting the tester or test user. If tests are validly used, in the traditional way, then the people who score highest on the selection test are selected as those most likely to do well on the job. If most of those obtaining higher test scores are white, then very few black workers will be able to obtain the jobs in question, because of the effects of social and economic disadvantage, and a long history of discrimination. The effect of the valid use of the test will be to perpetuate the patterns of racial inequality. If an employer has a social goal of recruiting more black workers, a conventionally used test (however valid) will

prevent their recruitment. According to most of the selection fairness models, in order to obtain the best balance of accurate prediction from the tests and at the same time make progress toward racial equality in employment by selecting more blacks, the user has to implement dual selection standards for the racial groups which ensure that greater proportions of blacks are selected. In effect the formulated models of selection fairness are different ways of achieving the desired balance of optimal prediction of job performance with minimal social cost in terms of exclusion of blacks from employment, education or training.

The differences in the selection models are unlikely to have important implications for employers in Britain as the definition of indirect discrimination in the Race Relations Act 1976 requires that selection procedures or criteria which have adverse impact, and which are the subject of a complaint or investigation, should be justified by the employer on non-racial grounds. This almost certainly means that the test be job-related in some sense. By this interpretation the models propounded by Cleary[26] and by Einhorn and Bass,[27] are most appropriate. According to Cleary's model, optimal prediction and selection fairness are taken to be strictly equivalent, and applicants with the highest predicted performance on the job, using separate bases forprediction for the ethnic/and sex sub-groups, would be selected. Different equations may be necessary for sub-groups to ensure that the prediction is maximally efficient within each sub-group, as prediction based on the group as a whole might, on average, over or under-predict job performance for members of the sub-groups. The Einhorn and Bass model uses a different definition of bias but the practical effects are the same.

Alternatively, the definition of indirect discrimination could be read to mean that, given that a complainant has established that a selection procedure is unfair because of his ethnic origin, etc., the employer may only have to demonstrate some degree of job-relatedness and not maximum job-relatedness. In other words it might be sufficient to fairly assess applicants for a minimal level of job suitability using valid techniques, and then to choose from among that group on some other basis. This would be consistent with normal practice as the test scores would be only one element in the selection decision.

This would enable employers to select higher proportions of women or ethnic minorities than might otherwise be the case, who would still be suitable for the job. The main disadvantage is that if done on a basis of systematically choosing black applicants from among those who are minimally suitable, it could result in direct discrimination against white applicants. One solution would be to choose, from among those applicants whom the tests reveal as minimally suitable, in such numbers that the

overall proportions selected from the applicant pool are the same for each ethnic group, but the difficulty here is that the decision would have to be based on non-racial grounds. In this way an employer could validly use tests to help select suitable employees, in such a way that indirect discrimination is avoided. The tests have the advantage of objectively appraising prospective employees and thereby reducing the risk of unfair discrimination and the costs resulting from inefficient selection. The employer gains compliance with the law, and makes progress toward genuine equality of opportunity, within the framework of an efficient process. This strategy has been used by a number of American employers who wish to use valid tests, but at the same time wish to recruit ethnic minority workers as part of an equal employment opportunity policy.

Perhaps it should be made clear here that the operation of racial minority quotas for employment selection is unlawful in the United States, just as it is in Britain. Some confusion has arisen over this issue because US courts can order quota hiring arrangements to compensate for past discrimination where a complaint against an employer has been upheld, but these are strictly temporary measures. Alternatively an employer may operate quotas by a voluntary agreement with the EEOC as part of a conciliated settlement. This agreement, or consent decree, is voluntarily entered into though it is enforceable through the courts.

One of the issues before the Supreme Court (October 1977) in the Bakke case is whether it is constitutional for an institution to voluntarily operate what in effect is a quota arrangement, even though a court has not judged that it had discriminated in the past*. However, goals and targets are lawful under American law. The claimed difference is that quotas by definition must be achieved (by selecting the right number of blacks), and failure to do so can result in the application of court sanctions.

* The controversial Bakke case went before Supreme Court in October 1977. Bakke, a white man, claims under the 14th Amendment of the Constitution, that he was denied equal protection of the law in his attempts to gain admission to a University of California Medical School because of his race. This is because the University of California operated a special admission Scheme which in two successive years admitted 16 disadvantaged black students all of whom had inferior school records and lower aptitude test scores than Bakke. A key issue at stake is whether it is constitutional to take account of someone's race in deciding admissions. (The case is now before the United States Supreme Court and the ruling will be of enormous importance to educational institutions and to employers who have programmes designed to recruit and select more black workers. The Court gave its ruling on 28th June 1978 and ruled that it is not unlawful to take race into account in selecting from applicants for higher education, but that the way in which the Medical School operated its "minority sensitive" Admissions Scheme is unlawful because it resulted in unfair discrimination against whites. The Court ruled that Bakke should be admitted, but gave no guidance on how an admissions scheme could be racially sensitive but not unfairly discriminatory against whites. This must await the arrival of similar cases already on their way to the Supreme Court.

By contrast, goals are objectives towards which the employer strives, and there may be acceptable reasons why achievement falls short of the target, or why progress is slow. In practice however the difference between goals and quotas are not so clear-cut, and many employers are prepared to hint or state boldly that they are operating quotas under the guise of goals and targets, as this was the best way to keep the EEOC off their backs. Federal contractors are required by the Department of Labor to set targets and goals, and to provide evidence of progress towards attaining them. In practice, this can mean the virtual imposition of quota hiring.

One way in which the different models of selection fairness could be used under the Race Relations Act 1976, is to regard them as operational definitions of different ways of selecting from among ethnic groups for the purpose of providing training for members of ethnic minority groups who are under-represented in particular jobs. (See sections 37-38 of the Act). The training should equip them with necessary skills to do those jobs so that they would then be better placed to compete for them in the labour market. Ideally the provider of such training should select those most likely to benefit from the training. A high success rate on completion of the training is important not only to ensure best use of resources, but also to encourage others to avail of the opportunity. The models could be used to select the desired proportions of ethnic minority groups, who would also have the necessary capacity for the training.

Whose rights: groups or individuals?

The special training provisions of the Race Relations Act are based on the assumption that disadvantage is equated, and in some sense identified, with racial or ethnic origin. The Act does not encourage the provision of similar programmes for disadvantaged whites who may live in the same declining inner city areas, though of course, there is no law to prevent it. The training provisions of the Act, and the models of selection discussed above, are based on a concept of fairness to groups, rather than to individuals, which is generally referred to as the group-parity model. This model has been examined on legal, social, and logical grounds by Novick and Ellis who find it unsatisfactory on several counts.[28] The argument is complex, and some of the issues are only relevant to the American Constitution. However the social limitations of the group-parity models are directly relevant to Britain. Novick has argued convincingly that the existing selection fairness models, and the group-parity concept on which they are based, are unsatisfactory because :

1) They often do not provide the amount of compensation for individual disadvantage that many people believe to be appropriate;*

2) The solutions which are implemented may increase advantage for some disadvantaged persons, but often they do so only at the expense of others who are more disadvantaged;

3) The solutions do not award a higher degree of advantage, or compensation, to those who are relatively more disadvantaged.

Novick argues for a new position, based on statistical decision theory, which emphasises the concept of utility**, and fairness to individuals rather than groups. This approach would treat individuals as individuals, providing compensation in direct proportion to their degree of disadvantage, and would not use racial group membership as a blanket index of disadvantage. Novick argues that this approach has many advantages. Primarily, it would command greater acceptability as it obviates the need for racially based preferences in selection and the quota hiring of blacks. In addition, for any given individual, a system of preference based on disadvantage is necessarily temporary. Once the disadvantage is alleviated then the selection preference no longer operates, whilst a system based on racial characteristics can operate indefinitely giving preference to those already free of disadvantage and may discourage attempts by the individual to obtain better education or training.

There are many technical problems associated with this approach, but perhaps the most important is the development of standard techniques for measuring disadvantage, which would be comparable to standard measures of mathematical or verbal ability, and which could be used as standard measures of mathematical or verbal ability, and which could be used as standard variables in the decision-making process.

* This point is particularly relevant to the situation in America, and refers specifically to graduate selection programmes which attempt to increase the selection of the disadvantaged, usually defined by reference to racial group membership. One practical application of selection fairness models to statistics from the University of Washington which was the subject of a Supreme Court case (Defunis v Odegaard) did not have the desired effect of recruiting as many black students as the "quota" system operated.

** The concept of utility takes into account the desirability of the positive or negative outcomes of the use of the test. These will differ according to whether a white or black person is concerned, in the context of the general goal of American society to achieve equal opportunity.

The first step involves arriving at a deeper understanding of the nature of individual disadvantage. Novick postulates a variety of factors which contribute to disadvantage including deprivation of exposure to 'educated' persons, economic deprivation, or unstable home environment, and also more specific factors, e.g. family size, birth order. It is clear that a considerable amount of research is required. Ten years ago the United States Employment Service* developed a measure of cultural exposure to determine the extent to which individual disadvantage affected performance on their own battery of aptitude tests. The measure was based on father's occupation, parents' education, public assistance to the family, type of guardian, family socio-economic status. These items were selected on the basis of showing good differentiation between school children in schools serving advantaged and disadvantaged metropolitan areas. Although the measure of cultural exposure correlated with performance on the aptitude tests, the predictive efficiency of the test scores was not improved when the individual's degree of cultural exposure was used in the prediction equation. As a result the research was abandoned. A standard scale of disadvantage will not be easy to achieve, but it would certainly be a valuable tool which would help circumvent the difficult moral issues of racially based preference in selection which are posed in the Bakke case.

A great deal of research is necessary before a standard measure of individual disadvantage can be developed. The definition of individual disadvantage would of necessity have to go beyond the rather limited demographic variables listed by Novick to include, in the case of blacks, low self-esteem, social alienation and the possibility of other psychological effects, resulting from systemic and direct discrimination against blacks from the time of slavery to the present day. In the meantime the group-parity concept which identifies disadvantage with racial group-membership, can be defended as administratively convenient. The criterion of 'race' for selection purposes is efficient if the aim is to recruit more blacks to compensate for discrimination in the past. An argument that is frequently put in the United States is that society cannot be colour-blind now if it wants to be colour-blind in the future: racially-based preferences in selection are necessary if the United States is to move from being a discriminatory society to a non-discriminatory society.

The outcome of the Bakke case, unless the Supreme Court evades the critical issue as it did in an earlier similar case (DeFunis v Odegaard), will be a decision whether

* Equivalent in Britain to the Employment Service Division of the Manpower Services Commission.

or not some discrimination in favour of blacks now is necessary if the goal of equality is to be attained in the future. In Britain, the Race Relations Act allows preferential training but not preferential selection. A key question in this context is whether objective tests should be encouraged or discouraged in view of the problems discussed above.

It is clear that many testing practices in the United States were not justifiable, and that one of the initial effects of the Civil Rights Act was to expose and to eliminate widespread bad testing practices. In addition it has stimulated new research, and the re-examination of old ideas. As a Washington lobby of employers, who were concerned at government intervention on the use of tests, put it:

> " ... federal guidelines on the validation of selection procedures have revitalised industrial and organisational psychology which over the last twenty-five years has been more complacent and pedantic than creative and critical.
>
> " ... many times more validation, and related research has been conducted in the last five years than in the previous twenty-five ..."[30]

Not only is more validation research being conducted, but there is also renewed debate about old ideas such as the nature test validity.[31] New kinds of validation strategy are being developed and assessed which attempt to overcome some of the limitations of the traditional strategies. One difficulty lies in the great specificity of validation studies, making it difficult to use the data obtained in one factory or department, in another even within the same organisation. A related area of concern is the stipulation of the conditions which permit the transportability of validity data from one situation to another. The practicalities of multi-unit or co-operative studies on a nation- or industry-wide basis are also being investigated. If these techniques are effective, more rather than less, validation can be conducted, with the result that the effectiveness of tests (and other selection criteria) can be established. This has led to a resurgence of interest in testing, by small as well as larger companies, since a validation study also indicates the increase in selection efficiency, so that the benefit or savings to the employer can be calculated.

The work of Melanie Baehr at the University of Chicago's Industrial Relations Centre is a good example of practical research with real benefits to the employer, as well as to prospective employees in terms of the prevention of racial discrimination.[32] She has conducted many validation studies in which she has tested the feasibility and meaningfulness of a variety of techniques. The results have also greatly clarified the legal standards which it is reasonable to impose on employers who use tests. In addition, she has argued that an effective strategy is to develop measures for use as selection devices which do not result in average differences in performance between ethnic groups.[33]

Another important development is the growing importance of formal techniques for the analysis of jobs in terms of the skills, knowledge and attributes required for successful performance.[34] Job analysis is emerging as a critical technique by which jobs can be compared (to permit generalisation of validity data from one context to another), to determine and develop selection criteria, and to form the basis of establishing the job-relatedness of selection procedures in validation studies, where a comprehensive objective description of the job is essential.

Testing in the United States is now on a newer, firmer footing, and should not be judged by the inadequacies of the past. Validly used job-related tests which take count of possible sources of bias can help eliminate arbitrary (often traditional) selection criteria and reduce the subjectivity of selection procedures, which may have large discriminatory effects. In this way an employer can obtain objective and relevant information about applicants, which should reduce the risk of unfair discrimination in selection. However, in contrast to whites, higher proportions of black workers may do less well on the tests and as a result fail to obtain jobs. By replacing arbitrary or inappropriate criteria, and supplementing subjective interviews with objective tests, some unjustifiable barriers to the employment of black workers can be removed, but the new standards, though job-related may still screen out proportionately more black workers, even with the safeguards against unfair assessment which have been outlined above. In this case it can more securely be argued that the rejected applicants are, on average, not suitable for the job as there is always a quantifiable margin of error. The benefit of this approach vis a vis the use of quotas, is that a realistic index can be obtained of the number of black workers who lack needed job skills, so that compensatory skill training can be provided, either by the employer, or by the Government. Consequently the more widespread use of tests will increase selection efficiency, remove unfair discriminatory barriers, and will also define the extent of the need for compensatory training of black workers.

It has already been pointed out that the use of quotas (or discrimination in reverse) is unlawful under the Race Relations Act. Even if it were not, arguments were put forward that reverse discrimination on the basis of race would not be desirable. The arguments for the use of quotas for blacks in employment and education in the United States are stronger than they are in Britain. The pattern of racism in the United States is more extreme, and spans a period of time which stretches back beyond the War of Independence. Consequently more extreme counter-measures may be necessary. In Britain the coloured population is seen (wrongly) as largely an immigrant population of fairly recent arrival. The restriction of coloured immigration is a key public and political issue. If this is the popular image it is doubtful how far reverse discrimination (or similar measures), even if lawful, could be acceptable in the public eye. This is an important consideration as adverse public reaction could easily jeopardise other areas of progress in the move toward racial equality. The more widespread use of valid tests based on objective job analyses, with due regard to the elimination of unfair bias in their application, coupled with intensive training for under-represented groups can be seen as a necessary experiment to eliminate discrimination in employment selection. Only, if and when, the experiment fails should other measures such as dual standards and quotas be considered.

It can be argued that the use of quotas and dual standards is not the only alternative to validly used objective tests. Other alternatives include lowering selection standards for all and compensating with increased on-the-job or pre-job training; the introduction of internships for people with low attainment but high potential; the restructuring of jobs in order to eliminate selection standards which have a discriminatory effect on black workers, etc. Many of these alternatives are feasible if employers and society as a whole are prepared to undertake the considerable re-organisation and restructuring that they would necessiate. They are certainly viable strategies, but really belong to a period when it is recognised that more immediate shorter-term solutions have failed. The problem of racial inequality is with us now, and immediate practicable solutions are required. It may be possible in the future, if the general climate of opinion has changed, to introduce the more fundamental, structural changes to bring about equal employment opportunity. In the meantime the culturally fair use of job-related tests may be the best means by which discriminatory barriers may be removed; and large-scale special training programmes the means by which equality of opportunity can be achieved.

II Guidelines on selection

Most of the issues and topics discussed above are not of merely academic or theoretical interest, as they have practical implications for United States employers who are all subject to both State and Federal anti-discrimination legislation. As noted earlier there are three Federal agencies which have responsibility for ensuring compliance, and all three have produced advisory or regulatory documents on selection procedures. In 1966 the EEOC published the first Guidelines on Employment Testing Procedures, which were designed to be a manual for personnel psychologists in their efforts to prevent discrimination. The 1966 Guidelines laid down generalised professional requirements. In particular they encouraged the use of job-related ability tests, and recommended that tests be judged in relation to whatever abstract or concrete attributes they were supposed to measure. In 1968 the Department of Labor, in its role of enforcing the Federal Contractor Compliance Program, issued a set of regulations on personnel testing which was revised in 1971 and 1974, and its coverage extended to other selection procedures. Although the general purpose of the Department of Labor's regulations was similar to the advisory guidelines issued by the EEOC, compliance with them was obligatory if an employer wished to obtain, or maintain, a government contract. Finally, the Federal Government, itself the largest employer in the United States with three and a half million employees, policed its own equal opportunity effort through the Civil Service Commission which issued a third set of Guidelines. This division of responsibility has caused confusion as well as conflict over the standards imposed.

For example, the employment practices of state and local government are regulated by the EEOC and the Civil Service Commission, and if Federal Contracts are involved, by the Department of Labor as well. Each organisation enforces compliance using different guidelines or regulations, and under different legal authority. It should have come as no surprise that the three organisations would have difficulty in agreeing with each other.

The conflict became critical in 1970 when the EEOC revised its Selection Guidelines, setting what many people believed to be impossibly high standards. By comparison with the earlier set, the 1970 Guidelines took a much tougher stand, setting out very detailed requirements. Many State Courts, and the two decisions in the Supreme Court discussed earlier, have in the eyes of many given the stringent EEOC Guidelines the force of law, even though this was not intended nor authorised by Congress, and the Guidelines were not subjected to public debate prior to their publication.

It soon became clear that the 1970 Guidelines are almost impossible to comply with completely, and there is indeed a conspicuous paucity of validation studies which have done so. The EEOC Guidelines also brought about an unacceptable state of affairs in which the Federal Government, through the EEOC, was applying stricter standards to private employers than to itself as an employer, through the Civil Service Commission. Finally, there was considerable confusion concerning which standards actually applied to a given organisation. This was nicely illustrated in the case of Douglas v. Hampton, which involved testing by the Civil Service Commission. In this case attorneys from two divisions of the Department of Justice were in court arguing on opposite sides of the same case, and the EEOC filed a brief in opposition to the Civil Service Commission.*

Since 1972 the major enforcement agencies represented on the specially formed Equal Employment Opportunity Co-ordinating Council (EEOCC)** have been attempting to reach agreement on a uniform set of Federal Guidelines on selection. After innumerable drafts of Uniform Guidelines, the Co-ordinating Council came close to agreement in 1976, but the EEOC, apparently after a dramatic 11th hour intervention by an influential civil rights group,*** withdrew at the last moment. Consequently the Departments of Labor and Justice and the Civil Service Commission jointly published the Federal Executive Agency (FEA) Guidelines in November 1976,[35] and the EEOC republished its own controversial Guidelines which had first appeared in 1970.[36] Several more drafts were prepared and rejected in the course of 1977, and a new draft was published in October.

* This case is interesting in that it involved the demonstration of the job-relatedness of the Federal Service Entrance Examination (FSEE) which followed the demonstration of a high degree of adverse impact on black applicants. The Civil Service Commission defended the tests on rational grounds. An Appeals Court declared that this is only acceptable where it can be demonstrated that empirical proof of validity for the test is infeasible. The CSC ceased operating the test, in favour of a new battery, the Professional and Administrative Career Examination (PACE), and so the injunction to prevent further use of the FSEE was declared moot. The Civil Service Commission has however accepted that criterion-related validity should be carried out if it is technically feasible, but has stressed that the cost would be enormous.

** Neatly referred to as the E^2OC^2.

*** The Legal Defence Fund of the National Association for the Advancement of Coloured People is a very influential pressure group, particularly with Congress, on many issues including testing.

The debate continues, not only because the issues are technically complex, but because there are many other aspects involved. A key figure in the debate recently drew attention to the "Legal, Political, Social, Technical, Professional, Ethical, and Emotional Aspects of Federal Guidelines on Employee Selection Procedures".[37] It is of value to examine some of these aspects as they might apply to Britain.

It has been argued by some that the main objective of the EEOC guidelines is to do away with, or at least severely restrict, testing by imposing stringent demands in terms of validation requirements whenever the smallest adverse impact is revealed. It has been claimed that the first effect of the Civil Right Act was to eliminate unprofessional practice in the use of tests, but now the more recent guidelines are discoureging the good users as well. Alternatively, it is argued that to use less stringent standards would permit test abuse and discrimination against blacks on a large scale. The civil rights groups have described the FEA Guidelines as a sell-out to the employers' interests.

To impose strict unrealistic standards on employers is to encourage them to do two things. Either they continue to use tests in a flexible way to avoid adverse impact, or they drop tests altogether, introducing non-quantitative methods which can also be used to avoid adverse impact. Some of these techniques, e.g. interviews, are almost impossible to validate, but at least they can appear to be fair to all ethnic groups, regardless of their inefficiency as a means of selecting the most suitable people for employment.

This has prompted R.M. Guion, a leading expert on testing, to remark cynically that under the law; "It's O.K. to be fairly stupid, but not O.K. to be unfairly stupid". In other words it doesn't matter how inefficient selection procedures are so long as there is no adverse impact on ethnic minorities. If on the other hand, an employer uses procedures, which have some adverse impact, which triggers the need to demonstrate validity, it is virtually impossible to satisfy the validity requirements as defined in the EEOC Guidelines. Consequently the employer is motivated to use techniques which, regardless of validity, ensure that appropriate numbers of black workers are recruited. Paradoxically, the EEOC seems to be achieving the desired objective (proportionate representation of black workers) by forcing employers to use <u>means</u> which are technically unlawful.

It would appear that, as far as the EEOC is concerned, compliance means no adverse impact, whereas to the employer, according to the law, it should mean that where there is adverse impact the employer should use valid procedures. If an employer can develop alternative procedures with equal validity but less adverse impact, then he

should use these instead. The objective should be the use of techniques which combine maximum validity with minimal adverse impact. The crux of the issue is whether the ultimate objective is to ensure that employers arbitrarily select larger numbers of ethnic minorities so that there is no adverse impact in selection, or whether the objective is genuine equality of opportunity.

Another issue with parallels in Britain, is the fact that the U.S. Civil Service Commission is obliged under its statutory authority to select solely on merit. In fact, one of the reasons why tests were originally introduced was to put selection on an open and objective basis, in order to safeguard the system from political and class favouritism, by selecting the best among the applicants as measured in part by the Commission's own tests. There has been little attempt to justify selection on the basis of the demonstrated job relatedness of the tests used. Instead the tests have been defended on the basis of rationalised or "rational validity", though it has been argued that the traditional interpretation of merit is too narrow, and that job relatedness should be an important component of selection by merit.[38] If as now seems the case the Civil Service is to demonstrate the validity and fairness of its tests, the cost of the exercise will run into millions of dollars. In Britain this raises the issue of the demonstration of the validity of Civil Service Selection schemes. A recently published study used, as a criterion, the career progress of those selected by the test battery over a period of 30 years.[39] This adequately demonstrated the utility of the test battery to the employer but in no way establishes the fairness of the selection, or indeed that of the promotional, system to the ethnic minorities. This is a case in which the promotional system itself might be unfair to ethnic (or sex) groups, so that any validity study based on it would not reveal unfair bias in the selection stage.

Another difficulty in the formulation of the Guidelines has been the need to lay down practical and attainable standards, whilst at the same time avoiding loopholes and loosely defined requirements which would allow evasion and misuse of tests and other selection devices. Another balance to be achieved is between setting minimal, practically feasible standards on the one hand, and ideals on the other. In the FEA Guidelines reference is made to the desirability of conforming to the professional standards embodied in the Standards of the American Psychological Association,[40] but these are really ideals which were not formulated with the view that they be given the force of law. As Novick has recently argued, this would certainly require a re-examination and revision of the 'standards' if they are to be used in this way.[41]

The standards embodied in the Guidelines are also causing difficulty, and in particular the exactness with which some of the provisions are expressed. The FEA Guidelines attempted to introduce rules-of-thumb so that employers would at least know the standards to which they were being held. A much discussed instance is the famous "80%" rule to determine adverse impact: If the selection rate for one group, e.g. Asians, is less than 80% of that for the majority group, then adverse impact is judged to be sufficiently large to require the demonstration of the job-relatedness of the selection device, as laid down in the Guidelines. In the absence of such a rule, despite its limitations, an employer can never know where he stands. This seems to have been the case with the EEOC Guidelines where any differences in selection rates, however small, could be taken as evidence of adverse impact, which triggers the requirement to show job-relatedness.

Many of the issues and techniques covered by the Guidelines are evolving and still controversial. A criticism that has been levelled at the EEOC Guidelines is that they impose standards which do not reflect the "state of the art" as reflected in the judgement of the American Psychological Association.

To some extent written selection tests have been made the scapegoat for discrimination that arises elsewhere in the selection process. This is partly because it is quantitative and therefore more amenable to examination for discriminatory effect. If the validation requirements embodied in the Guidelines are set at unrealistic levels, do not allow for innovation, and do not reflect the current (evolving) state of the art, then the use of the tests will be discouraged, and discriminatory effect will be concealed within a host of subjective factors in selection. In addition some employers may be encouraged to operate unofficial quotas.

The eventual Guidelines must inevitably be a compromise and contain reasonable practical guidance for practioners, which set attainable standards, and which are written with sufficient precision for clear interpretation by employers as well as in law. The effort to produce uniform Guidelines has been an enormous power struggle involving psychologists and lawyers, test publishers, civil rights groups, employers and the different federal enforcement agencies, and it may not yet be over. Although the amount of time invested (nearly 6 years) may be great, it has at least had the advantage of generating real debate, the clarification of legal obligations, the re-examination of old ideas (especially test validation), and has provided the stimulus for research into new techniques, and the applicability of specific provisions of the Guidelines. In the long-term, and it is easy to understand those who are impatient, this can only lead to an improvement in the effort toward genuine equal opportunity, and away from a society in which the rights of some can be upheld only by taking the same rights away from others.

The six years of wrangling between the Federal Agencies over the Selection Guidelines has yielded some practical lessons for Britain, and the formulation of Codes of Practice to be issued by the Commission for Racial Equality, under Section 47 of the Race Relations Act. The Code is not legally binding as its function is advisory. However an industrial tribunal may consider the extent to which an employer has complied with the Code, and can be taken into account in deciding a particular question. The Code is therefore a critical document in interpreting the law for the guidance of individuals, employers, and the tribunals.
Selection for jobs or promotion is a critical area to be covered in the Code. The lessons that can be drawn from American experience are many :

(a) It is important to draw up a Code that represents the best possible balance of all the interested parties, which in addition to the ethnic minorities include private, and public employers, local and central government as employers, test distributors, and the professional standards of occupational psychologists.

(b) The Code should be as practical as possible, and should express more than a broad general intention to avoid discrimination in selection. The process is far too complex, and the consequences too important, to leave to the mere expression of intentions.

(c) The Code should be written on the understanding that the employer and other test users need to know exactly where they stand under the law, or failing this, at least to have a more precise idea than the vague statutory definition of indirect discrimination allows.

(d) The Code should give employers some indications of what would constitute a demonstration of how job requirements can be justified on non-racial grounds. There should be an attempt to define the job-relatedness of a selection procedure.

(e) It would be unwise to rely on the assumption that all relevant professional standards are sufficiently clear or agreed to resolve difficult areas not explicitly touched on in the Code.

(f) The Code should be written on the understanding that it may be used by industrial tribunals to judge complex technical matters in which they are not themselves expert.

(g) It may be preferable to utilise some simple rules of thumb.

III Conclusions

The impact of the Civil Rights Act 1964 on testing practices in the United States has been profound. The early days of the legislation saw the exposure and outlawing of gross malpractice. A panic phase seems to have followed during which there was a feeling that even good testing practice could not be defended before the courts. This led to a dramatic fall in the use of employment tests. To some extent this fear still persists as many employers who formerly used tests on a large scale still refrain from doing so. In recent years however there has been a resurgence of interest in testing, which is undoubtedly on a surer footing than it was in previous years. This resulted from the generation of new debates about old ideas, the investigation of test fairness, and the extensive theoretical development coupled with innovative research which has been undertaken, particularly in the last ten years. In addition the controversy over testing, and the Federal Enforcement Agencies Guidelines has brought to the fore many social, moral, and technical issues. Many of these have direct relevance to Britain.

Employment testing is still not a widespread practice in Britain, and there is a great potential for growth in the use of tests, for there can be no doubt that properly used, formal quantified assessment devices can yield considerable benefits to the employer and to the employee. Awareness of sex or racial bias in testing should not discourage employers from using tests. Indeed, with due regard to the reduction of bias in their use, validly used job-related tests can contribute to the removal of arbitrary barriers to the employment of black workers.

It is unlikely that in the foreseeable future pressures will be brought to bear on British employers which are in any way as extreme as those operating on American employers to create equal opportunity for minorities. In the absence of powerful civil rights groups, or the deterent value of severe court penalties, or a forceful energetic law enforcement agency, the climate in Britain is very different. Even so there are some basic points which have general applications.

1. Employers should examine the impact of their selection procedures on minority groups to determine whether disproportionately larger numbers of ethnic minority groups are being rejected.
2. If there is an adverse impact, the employer should ensure that the practice or standard causing the impact is justifiable in terms of the job involved. This requires a validation study. (Even if there is no adverse impact it makes sound commercial sense to use valid procedures for selecting staff).
3. It is impossible to detect adverse impact if the ethnic origin of applicants and existing employees is not recorded in some form.
4. If the selection procedure can be justified on non-racial grounds (i.e. is valid), then the employer is within the law, despite any adverse impact. If he is concerned at rejecting large numbers of ethnic minorities then he can seek procedures which are equally valid, but less discriminatory.
5. Objective job analysis techniques are an indispensible tool for setting selection standards, and validating selection procedures. They would form the most objective and reliable basis for demonstrating that a procedure is justified on non-racial grounds.
6. If ethnic minority groups tend to obtain lower scores, this is not a sign that the test is unfair. It depends entirely on the interpretation put on the test and the general way in which it is used.

Above all the American experience has shown, despite its controversies, soul-searching, and confusion that a recognition of the difficulties in the road is not a reason for giving up, but is a reason for being more careful, and for exploring new routes. In Britain, one route worth exploring is the wider use of job-related selection tests, which can be used to eliminate or circumvent discriminatory barriers to the employment of black workers.

References and Notes

1. Sneath, F. and Thakur M. and Medjuck, B. (1976) Testing People at Work. Information Paper 24. London : Institute of Personnel Management.

2. Kamin, L.J. (1977) The Science and Politics of IQ. Harmondsworth : Penguin.

3. The debate was re-opened by Arthur Jensen's article: "How much can we boost IQ and scholastic achievement?" Harvard Business Review, 39, 1,1-123.

4. Race Relations Act 1976, Section 1 (1) (b).

5. Racial Discrimination. Cmnd 6234. September 1975. London : H.M.S.O.

6. Executive Orders from President Johnson (No. 11246 in 1965, and No.11375 in 1967) required Federal contractors to take positive action to ensure not only that they did not discriminate but that they also provided equal employment opportunities.

7. Equal Employment Opportunity : Programs and Results. PPF Survey No.112 March 1976. Washington D.C. : Bureau of National Affairs.

8. See Carby, K. and Thakur, M. (1977) No Problems Here? : Management and the Multi-Racial Workforce. London : Institute of Personnel Management.

9. See Rees, T. (1974) Equal Employment Opportunity in the USA - Recent Developments. London : The Runnymede Trust.

10. The Employment of Federal EEO Mandates : A National Challenge. July 1977. Washington D.C. : National Urban League.

11. The American Underclass. Time Magazine, August 26, 1977.

12. The Concept of "The Underclass" was developed by the Swedish economist Gunner Myrdal.

13. The papers by Thaddeus Hot are "Personnel selection and the Supreme Court" and "A view from Albermarle". The latter appeared in Personnel Psychology, Vol 30,1, 65-80.

14. The Supreme Court case of Griggs v. Duke Power Co. is described at length in Pollak, L.H. (1974) Discrimination in Employment : The American Response. London : The Runnymede Trust.

15 In addition to the 1975-76 Compendium prepared by Dee-Ann Horstman, another guide examines cases as far back as 1971. Equal Employment Opportunity Court

Cases. Washington D.C. : U.S. Civil Service Commission.

16. In particular see Flaugher, R.L. (1974) Bias in Testing : A Review and Discussion, TM Report 36; and (1977) The Many Definitions of Test Bias, RM-77-2. Princeton, N.J. : Educational Testing Service.

17. Ebel, R.C. (1976). In defence of standardised testing. Paper presented at the Houghton Mifflin Annual Measurement Conference. March 31st, 1976, Denver City, Colorado. Quoted in Flaugher (1977).

18. See, for example, Kirkpatrick, J.J., Ewen, R.B., Barrett, R.S. and Katzell, R.A. (1969) Testing and Fair Employment. London : University of London Press.

19. Pearn, M.A. (1977) Selecting and Training Coloured Workers. Training Information Paper No. 9. Training Services Agency. London : H.M.S.O.

20. Williams, R.L. (1975) Black Intelligence Test of Cultural Homogeneity ; Manual of Directions. St. Louis, Missouri : Williams and Associates.

21. Arvey, P.D. (1972) Some comments on culture-fair tests. Personnel Psychology, 25, 433-443.

22. See the recent issue of Journal of Applied Psychology, Vol. 62,1977, which was entirely devoted to the debate. More recent issues continue the debate. See Vol. 63, February, 1978.

23. Campbell, J.T., Crooks, L.A., Mahoney, M.H., and Rock, D.A. (1973) An investigation of sources of bias in the prediction of job performance : A six year study. PR 73-37. Princeton, NJ : Educational Testing Service.

24 A complete issue of the Journal of Educational Measurement, Vol. 14,2, was devoted to Tailored Testing.

25. Petersen, N.S. and Novick, M.R. (1976) An evaluation of some models of culture-fair selection. Journal of Educational Measurement, 13, 1, 3-29. The whole of this issue was devoted to papers on culture-fair selection.

26. Cleary, T.A. (1968) Test bias : prediction of grades of Negroes and white students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

27. Einhorn, H.J. and Bass, A.R. (1971) Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 75,4, 261-269.

28. Novick, M.R. and Ellis, D.D. (1977) Equal opportunity in educational and employment selection. American Psychologist, May, 1977.

29. Droege, R.C. and Hawk, J. (1969) Development of a measure of cultural exposure. Personnel Psychology, 22,495-503.

30. Quoted in Schwartz, D.J. (1976) OFCCP : New Directions Under the Uniform Guidelines. Speech delivered to the Personnel Testing Council at Southern California Newport Beach, California, October. The Ad Hoc Industry Group is a powerful Washington lobby, which is very concerned at Federal intervention in selection procedures.

31. See for example Personnel Psychology, 1977, Vol. 30 which contains papers, by Guion, Tenopyr and others, from a conference on content validity.

32. Baehr, M.E. (1977) Impact of Civil Rights legislation and court actions on personnel procedures and practice. Chapter in "Police Selection".

33. Baehr, M.E. (1974) National Validation of a Selection Test Battery for Bus Operators : Executive Summary. Chicago, Ill.: University of Chicago Industrial Relations Centre.

34. Baehr, M.E. (1976) A Practitioner's View of EEOC Requirements With Special Reference to Job Analysis. Occasional Paper 37. Chicago, Ill. : University of Chicago Industrial Relations Centre.

35. The Federal Executive Agency Guidelines on Employee Selection Procedures, issued by the Department of Justice, Department of Labor and the U.S. Civil Service Commission, appeared in Federal Register, Vol. 41, No. 227, November 23, 1976.

36. The Guidelines on Employee Selection Procedures issued by the Equal Employment Opportunity Commission appeared in Federal Register, Vol, 35, August 1, 1970. They were subsequently re-published in the Register in November 1976, after the failure of the four agencies to agree on a set of Uniform Guidelines on Selection Procedures.

37. Gorham, W.A. (1977) Chapter Heading in Contemporary Personnel. Chicago, Ill. : St. Clair Press.

38. Bemis, S.E. (1977) Federal Government Intervention in Psychological Testing : Response from the Public Sector. Paper presented at the Annual Convention of

the American Psychological Association, San Francisco, August, 1977.

39. Anstey, E. (1977) A 30-year follow-up of the CSSB procedure, with lessons for the future. Journal of Occupational Psychology, 50. 149-159.

40. Standards for Educational and Psychological Tests prepared by the American Psychological Assocation, the American Educational Research Association, and the National Council on Measurement in Education. Washington, D.C. : American Psychological Assocation, 1974.

41. Novick, M.R. (1977) The Influence of the law on professional measurment standards. Paper presented at ETS Invitational Conference, October 29. Princeton, NJ : Educational Testing Service.